e3_c7-00

ATL-042

A

03/06/00 jc784 U.S. PTO

jc760 U.S. PTO 09/518937 03/06/00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Ajay Divakaran
Anthony Vetro
Huifang Sun

Title: METHOD FOR REPRENSENTING AND COMPARING MULTIMEDIA CONTENT ACCORDING TO RANK

"EXPRESS MAIL" mailing label number EJ690964877US

Date of Deposit: March 6, 2000

I hereby certify that this correspondence is being deposited with the United States Postal Service "EXPRESS MAIL Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231.

Kelli J. Withrow
Printed name of depositor

Kelli J. Withrow
Signature of depositor

09518937 030600

* * *

TRANSMITTAL OF FILING UNDER 37 CFR 1.53(b)

Assistant Commissioner for Patents
Washington, DC 20231

ATTN: Box Patent Application

Sir:

This is a request for filing a Continuation-in-Part (CIP) application under **37 CFR §1.53(b),** of pending prior application Serial No. 09/385,169, filed on August 30, 1999 entitled METHOD FOR REPRESENTING AND COMPARING MULTIMEDIA CONTENT

1.Fee Calculation (37 CFR 1.16)

CLAIMS AS FILED

	Number filed	Number extra	Rate	Calculations
Total Claims (37 CFR 1.16(c))	13 - 20 =	0	0 $ 18.00	$
Independent Claims (37 CFR 1.16(b))	1 - 3 =	>	X $ 78.00	$ >
Multiple Dependent Claim(s) if any (37 CFR 1.16(d))			+ $ 260.00	$ >
		Basic Fee		+ $ 690.00
	Total of above Calculations =			$
			Assignment	$ 40.00
			TOTAL	$ 730.00

2. Assignment

[X] an assignment of the invention to Mitsubishi Electric Information Technology Center America, Inc. is attached. A separate "ASSIGNMENT COVER LETTER ACCOMPANYING NEW PATENT APPLICATION" is also attached.

3. Fee Payment Being Made At This Time

[X] charge Account No. 50-0749 in the amount of $730.00. A duplicate of this request is attached.

4. Authorization To Charge Additional Fees

[X] The Assistant Commissioner for Patents is hereby authorized to charge the following additional fees which may be required by this paper and during the entire pendency of the application to Account No. 50-0749

[X] 37 CFR 1.16(a), (f) or (g) (filing fees)
[X] 37 CFR 1.16(b), (c) and (d) (presentation of extra fees)

5. Power of Attorney

[X] The power of attorney in the prior application is to Dirk Brinkman, Reg. No. 35,460.

[X] All future correspondence should be addressed to:

Patent Department
Mitsubishi Electric Information Technology
Center America, Inc. (ITA)
201 Broadway, 8th Floor
Cambridge, MA 02139

Respectfully submitted,

Mitsubishi Electric Information Technology
Center America, Inc.

By: _____
Dirk Brinkman
Reg. No. 35,460
Attorney for Assignee

Patent Department
Mitsubishi Electric Information Technology
Center America, Inc. (ITA)
201 Broadway, 8th Floor
Cambridge, MA 02139
(617) 621-7539

Method for Representing and Comparing Multimedia Content According to Rank

CROSS-REFERENCE TO RELATED APPLICATION

This is a Continuation-in Part Application of U.S. Patent Application of U.S. Patent Application Serial number 09/385,169, "Method for Representing and Comparing Multimedia Content" filed on August 30, 1999 by Lin et.

FIELD OF THE INVENTION

This invention relates generally to processing multimedia content, and more particularly, to representing and comparing ranked multimedia content.

BACKGROUND OF THE INVENTION

There exist many standards for encoding and decoding multimedia content. The content can include audio signals in one dimension, images with two dimensions in space, video sequences with a third dimension in time, text, or combinations thereof. Numerous standards exist for audio and text.

For images, the best known standard is JPEG, and for video sequences, the most widely used standards include MPEG-1, MPEG-2 and H.263. These standards are relatively low-level specifications that primarily deal with the spatial compression in the case of images, and spatial and temporal compression for video sequences. As a common feature, these standards

perform compression on a frame basis. With these standards, one can achieve high compression ratios for a wide range of applications.

Newer video coding standards, such as MPEG-4, see "Information Technology -- Generic coding of audio/visual objects," ISO/IEC FDIS 14496-2 (MPEG4 Visual), Nov. 1998, allow arbitrary-shaped objects to be encoded and decoded as separate video object planes (VOP). This emerging standard is intended to enable multimedia applications, such as interactive video, where natural and synthetic materials are integrated, and where access is universal. For example, one might want to "cut-and-paste" a moving figure or object from one video to another. In this type of scenario, it is assumed that the objects in the multimedia content have been identified through some type of segmentation algorithm, see for example, U.S. Patent Application Sn. 09/326,750 "Method for Ordering Image Spaces to Search for Object Surfaces" filed on June 4, 1999 by Lin et al.

The most recent standardization effort taken on by the MPEG committee is that of MPEG-7, formally called "Multimedia Content Description Interface," see "MPEG-7 Context, Objectives and Technical Roadmap," ISO/IEC N2729, March 1999. Essentially, this standard plans to incorporate a set of descriptors and description schemes that can be used to describe various types of multimedia content. The descriptor and description schemes are associated with the content itself and allow for fast and efficient searching of material that is of interest to a particular user. It is important to note that this standard is not meant to replace previous coding standards. Rather, it builds on other standard

representations, especially MPEG-4, because the multimedia content can be decomposed into different objects and each object can be assigned a unique set of descriptors. Also, the standard is independent of the format in which the content is stored. MPEG-7 descriptors can be attached to compressed or uncompressed data.

Descriptors for multimedia content can be used in a number of ways, see for example "MPEG-7 Applications," ISO/IEC N2728, March 1999. Most interesting, for the purpose of the description below, are database search and retrieval applications. In a simple application environment, a user may specify some attributes of a particular object. At this low-level of representation, these attributes may include descriptors that describe the texture, motion and shape of the particular object. A method of representing and comparing shapes has been described in U.S. Patent Application No. 09/326,759 "Method for Ordering Image Spaces to Represent Object Shapes" filed on June 4, 1999 by Lin et al. One of the drawbacks of this type of descriptor is that it is not straightforward to effectively combine this feature of the object with other low-level features. Another problem with such low-level descriptors, in general, is that a high-level interpretation of the object or multimedia content is difficult to obtain. Hence, there is a limitation in the level of representation.

To overcome the drawbacks mentioned above and obtain a higher-level of representation, one may consider more elaborate description schemes that combine several low-level descriptors. In fact, these description schemes may

even contain other description schemes, see "MPEG-7 Description Schemes (V0.5)," ISO/IEC N2844, July 1999.

As shown in Figure 1a, a generic description scheme (DS) has been proposed to represent multimedia content. This generic audio-visual DS 100 includes a separate syntactic DS 101, and a separate semantic DS 102. The semantic structure refers to the physical and logical signal aspects of the content, while the semantic structure refers to the conceptual meaning of the content. For a video sequence, the syntactic elements may be related to the color, shape and motion of a particular object. On the other hand, the semantic elements may refer to information that cannot be extracted from low-level descriptors, such as the time and place of an event or the name of a person in the multimedia content. In addition to the separate syntactic and semantic DSs, a syntactic-semantic relation graph DS 103 has been proposed to link the syntactic and semantic DSs.

The major problem with such a scheme is that the relations and attributes specified by the syntactic and semantic DS are independent, and it is the burden of the relation graph DS to create a coherent and meaningful interpretation of the multimedia content. Furthermore, the DSs mentioned above are either tree-based or graph-based. Tree-based representations provide an efficient means of searching and comparing, but are limited in their expressive ability; the independent syntactic and semantic DSs are tree-based. In contrast, graph-based

representations provide a great deal of expressive ability, but are notoriously complex and prone to error for search and comparison.

For the task at hand, it is crucial that a representation scheme is not limited to how multimedia content is interpreted. The scheme should also provide an efficient means of comparison. From a human perspective, it is possible to interpret multimedia content in many ways; therefore, it is essential that any representation scheme allows multiple interpretations of the multimedia content. Although the independent syntactic and semantic DS, in conjunction with the relation graph DS, may allow multiple interpretations of multimedia content, it would not be efficient to perform comparisons.

As stated above, it is possible for a DS to contain other DSs. In the same way that the generic DS includes a syntactic DS, a semantic DS, and a syntactic/semantic relation graph DS. It has been proposed that the syntactic DS 101 includes a segment DS 105, a region DS 106, and a segment/region relation graph DS 107. As shown in Figure 1b, the segment and region DSs may be used to define the temporal and spatial tree structure of multimedia content, respectively, and the segment/region relation graph DS may be used to describe the spatio-temporal relationships between segments and regions. Similarly, as shown in Figure 1c, the semantic DS 102 includes an event DS 108, an object DS 109, and an event/object relation graph DS 110. The event and object DSs may be used to define event and object trees that define semantic index tables for temporal events and spatial objects, respectively. The event/object relation

graph DS may be used to describe any type of spatio-temporal relationship between events and objects. As with the higher level DSs, namely the semantic and syntactic DSs, these lower-level DSs suffer the same problems with expressiveness and computational complexity.

Therefore, there is a need for representing syntactic and semantic attributes of multimedia content that balances the complexities of data structures and the methods that operate on the structures. In addition, there is a need to compare multimedia content according to content attributes.

SUMMARY OF THE INVENTION

The present invention provides a new method of representing syntactic and semantic attributes of multimedia content. It is an object of the invention to use existing attributes that may be contained within a semantic or syntactic description scheme using a framework that balances the restrictions on structure and expressiveness of elements, with the computational complexity of operations on those elements.

The method according to the invention is based in part on directed acyclic graphs (DAG). It is well known that the DAG occupies a middle ground between tree-based and graph-based representations. In addition, the DAG provides a new functionality of composition. In other words, many structural

compositions of an entity can be described by many structural compositions of its contained elements.

Most importantly though, the similarity between these structural compositions and the structural compositions created by another entity can be easily computed. Within this composition framework, the DAG also provides a means of combining syntactic and semantic elements so that similarity comparisons may seamlessly switch between both types of descriptions. In some sense, this can be viewed as a unification between the syntactic and semantic parts of the description scheme.

The method for representing the semantic and syntactic elements in a unified way also provides a means for unifying the spatial and temporal elements of multimedia content. The invention relies on the fact that the compositions referred to earlier are spatio-temporal compositions that contain both syntactic and semantic elements. The important points to keep in mind is that the compositions according to the invention are DAG representations, which facilitate multiple interpretations and low complexity comparison, and that the compositions, which define spatio-temporal attributes, both syntactic and semantic, are contained within the respective content entities.

More particularly, the method generates a representation of multimedia content by first segmenting the multimedia content spatially and temporally to extract objects. Feature extraction is applied to the objects to produce semantic and

syntactic attributes, relations, and a containment set of content entities. The content entities are coded to produce directed acyclic graphs of the content entities. Edges of the directed acyclic graphs represent the content entities, and nodes represent breaks in the segmentation. Each directed acyclic graph represents a particular interpretation of the multimedia content.

In one aspect the multimedia content is a two dimensional image, and in another aspect the multimedia content is a three dimensional video sequence.

In a further aspect of the invention, representations for different multimedia contents are compared based on similarity scores obtained for the directed acyclic graphs. Attributes of each content entity are measured and the measured attributes are assigned to each corresponding content entity in the directed acyclic graphs to rank order the multimedia content.

In another aspect of the invention, attributes of each content entity are measured, and the entities are ranked according to the measured attributes. The rank list can be culled for desirable permutations of primary content entities as well as secondary entities associated with the primary entities. By culling desirable permutations, one can summarize, browse or traverse the multimedia content. For example, the most active and least active video segments of a video sequence form a summary that has the desirable attribute of conveying the dynamic range of action contained in the video sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1a-1c are block diagrams of prior art description schemes;

Figure 2 is a block diagram of a description scheme for a general content entity according to the invention;

Figures 3a-3c are a block diagrams of description schemes for example content entities;

Figure 4 is a flow diagram of a method for generating the description scheme according to the invention;

Figure 5 is a flow diagram for a method for comparing the description schemes according to the invention; and

Figure 6 is a block diagram of a client accessing multimedia on a server according to the invention.

Figure 7 is a ranked graph; and

Figure 8 is a summary of the graph of Figure 7.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Introduction

We describe methods for representing and comparing multimedia content according to a ranking of the content. The methods are based a new generic data structure, which includes a directed acyclic graph (DAG) representation. In the following, we describe objects in our scheme and the advantages of the DAG representation. It is the DAG representation that allows the scheme to infer multiple interpretations of multimedia content, yet still be efficient in the comparison with other multimedia content. In fact, when we score with respect to a probability likelihood function, the computations are not only tractable, but also optimal.

Besides describing the generic data structure, we also describe three important functions that allow us to realize this efficient representation and perform comparisons. The first function will be referred to as a DAG-Coder. The DAG-Coder is responsible for taking individual content entities contained in the object and producing a DAG-Composition. The second function is an Object-Compare. The Object-Compare efficiently compares two content entities by determining a similarity score. The third function is Content Ranker. This function ascribes a ranking score to content entities so that DAG-Compositions can be traversed, browsed, or summarized according to rank. The traversing, browsing, and summarizing can be an increasing or decreasing rank order.

After the data structure and three functions mentioned above have been described, we review and elaborate on applications that are enabled by our representation scheme. An integrated application system that performs feature extraction, database management and object comparison is described. Also described is an application system for traversing, browsing, and summarizing multimedia content according to a ranking of the content.

Generic Description Scheme of a Content Entity

To introduce our scheme of representing content objects, we define generic object types, and restrictions on instantiations of such generic object types.

As shown in Figure 2, a content entity, for example, a video entity 200 is the main part of our scheme. The content entity is a data object that relates contained objects together. The content entity is a recursive data structure divided into four parts: attributes (properties) 201, relations 202, DAG-Compositions 203, and a containment set 204.

Attributes

The attributes 201 form the basis within our recursive description scheme. Attributes are an unordered set that contains properties that may provide details about parts of the entity or summarize the entity as a whole. Attributes are global to the object and may refer to such syntactic properties as color and

motion, or other semantic properties of the object such as time and place. The attributes provide basic, low-level information without any structure, however, after structure is added, it is these properties that actually contribute to the degree of similarity. Also, as we will describe later, attributes can define an ordering that help to compose and interpret the individual entities contained within the object. It should be noted that these properties are inherent qualities of the content entity that contains them and instantiations of this entity should be accessible/visible through the content entity itself.

As an example, a video sequence of a airplane landing on a runway may contain the semantic attributes of place, date, time and temperature, along with the caption, "airplane (767) landing." Some syntactic attributes that may be attached to this multimedia content are the trajectory of descent. Attached to the airplane object may be the color and shape of the airplane itself. Here, we make an important distinction between attributes of the multimedia content and attributes of the objects. The reason that the trajectory is an attribute of the multimedia content is because trajectory is relative to the ground. Therefore, it does not make sense as an attribute of the plane alone, whereas color and shape do make sense.

Relations

The relations (R) 202 are objects that detail relationships between content entities (VE). It is important to note that the context of the relations is given by the containing content entity. The reason is that multimedia content that are

segmented differently will produce different relations. Essentially, the relation can be viewed as a hyperlink between a contained object and something else, for example, another content entity. Types of relations are global and instantiations of relations should only be accessible within the content entity itself. One of the utilities of relations is that they may be useful in guiding a search. Returning to our example of the airplane landing, several relations can be identified: the plane is landing on the runway, the lights are guiding the plane, and the runway is located at a particular airport with a particular orientation.

The relations are different from containment, described below, in that the related object may not be completely contained by the content entity and therefore is not considered in similarity comparisons. However, relations allow a user to search for other relevant objects to the content entity in question. All the relations in the content entity must have one argument that is contained within the content entity.

DAG-Compositions

In general, the DAG-Compositions 203 are directed acyclic graphs 205 where edges 206 represent content entities and nodes 207 correspond to breakpoints in the segmentation. The DAG-Composition allow us to infer multiple interpretations of the same multimedia content. Because DAGs operate on 1D spaces, segmentation in this context refers to the delineation of some 1D process. For instance, if we consider a spatio-temporal multimedia content, then

the temporal segmentation is a 1D process that defines points in time where several successive events may begin and end. Hence, we may have a DAG-Composition that corresponds to temporal actions. In the spatial domain, we may define an order from left to right across an image. In this way, we may have a DAG-Composition that corresponds to object positions from left to right. Of course, we may define other orderings such as a counter-clockwise spatial ordering, which may serve a totally different purpose.

In U.S. Patent Applications Sn. 09/326,750 and 09/326,759, incorporated herein by reference, Voronoi ordering functions were respectively defined over the exterior and interior image space with respect to an object boundary. The ordering on the interior space was particularly useful in obtaining a skeleton-like representation of the object shape, then forming a partially ordered tree (POT), which made use of the DAG representation.

It should be emphasized though that the method of ordering 2D images or 3D video sequences to achieve DAG-Compositions is not the focus here, rather we are concerned with techniques that use the DAG-Composition to infer higher-level interpretations of a particular multimedia content.

Containment Set

The containment set 204 includes pointers to other content entities that are strictly contained temporally and/or spatially within the content entity 200. The restriction on the containment set is that one object cannot contain another

object that contains the first object, i.e., containment induces a directed acyclic graph. The content entities need not be mutually exclusive and there is no ordering within the containment set. For example, in the video sequence of the airplane landing, the containment set includes pointers to each content entity. Some possibilities include pointers to the plane, the runway, the runway lights, the plane touching down, radio communications, etc.

DAG-Coder

The DAG-Compositions are the result of different DAG-Coders applied to the content entity. In other words, given the content entities in the containment set and their relations, different DAG-Coders produce different interpretations of the multimedia content. This function is further described in the following.

A DAG-Coder is a function that segments a given content entity into its components by inducing an ordering over the content entity components. The DAG-Coder produces the DAG-Composition 204. The DAG-Coder is global to the database and can be applied to any content entity. The DAG-Coder provides a perspective on the spatio-temporal content space and make similarity calculations between objects more tractable. A path in the DAG represents an interpretation of the content entity 200. This DAG representation becomes a framework for the description scheme that can interchange syntactic and semantic information, at any level. Furthermore, the complexity of the description scheme is hidden from the user.

Multiple Path through a DAG

The DAG-Coder produces multiple interpretations of the multimedia content through such DAG-Compositions. This is achieved through the multiple path structure of the DAG. In the following, we focus on what these multiple paths really mean in terms of the multimedia content.

Figures 3a-3c illustrate multiple paths in terms of an example "baseball video" entity 300. In Figure 3a, the content entity 300 includes attributes 301, relations 302, DAG-compositions 303, and a containment set 304. In Figure 3b, a content entity 310 includes attributes 311, relations 312, DAG-Compositions 313, and a containment set 314.

As illustrated, a temporal DAG can represent equivalent interpretations of the same event. For instance as shown in Figures 3a and 3b, in the baseball video, a pitching and hitting sequence, or the inning that is being played may be recognizable through the observation of syntactic elements, such as motion, color and/or activity. However, as an alternate means of representation as shown in Figure 3c, such a sequence or event can also be summarized by attributes 321 of the commentary of the announcer 320. So, from this example, is evident that multiple temporal interpretations of multimedia content are possible and that they may or may not occur simultaneously.

In the case of spatial DAGs, multiple paths can also represent equivalent interpretations, and in some sense can add a higher level of expressiveness. This added level of expressiveness is achieved by a grouping of individual objects into a composite object, then realizing that this composite object can be interpreted with a different semantic meaning. Usually, this new semantic interpretation is higher than before since more information is considered as a whole.

As an example, consider several objects: a gasoline pump, a gas attendant and a car. Individually, these objects have their own set of attributes and are distinct in their semantic meaning. Put together though, these individual objects can obviously be interpreted as a gas station. These multiple paths are efficiently represented by the DAG structure. On the syntactic side, various interpretations of the shape of an object for example may be deduced in a similar manner.

Generating Multimedia content Description

Fig. 4 illustrates a method 400 for generating a description scheme 409 from a multimedia content 401. The multimedia content can be a 2D image or a 3D video sequence. First, spatial and temporal segmentation 410 is applied to the multimedia content to extract objects 411. Next, feature extraction 420 is applied to the objects to obtain a set of all content entities 429. Feature extraction includes attribute extraction 421, containment extraction 422, and relations extractions 423. The DAG-Coder 430, according to an ordering 431,

generates the DAG-Compositions for the entities 429 to form the multimedia content description 409 according to the invention.

Comparing Different Multimedia Content

Figure 5 shows a method for comparing two different multimedia contents, content 1 501 and content 2 502. The method generates 400 two description schemes, DS_1 503 and DS_2 504. The descriptions are compared 510 to produce a similarity score 509. Given two types of objects, the object comparator returns a similarity score in terms of the probabilistic likelihood that the two objects are the same. The Object-Compare function 510 may recursively call other Object-Compare functions. The Object-Compare is very similar to the algorithm using for comparing Partially Ordered Trees (POT) as described in U.S. Patent Application Sn. 09/326,759 incorporated herein by reference. The key points are reviewed below.

We consider the matching algorithms used to compare Ordered Trees. Because trees are recursive structures, we can do optimal comparisons recursively and base the comparisons upon single node trees. Let us score our trees in the range of 0 to 1. Two single node trees are assigned a score of 1, while a single node tree and any tree with more than one node is assigned a score of 0.

For our inductive step, we note that each node corresponds to a sequence of edges and their respective children. To compare trees, we merely find the best

correspondence between the sequences, while recursively comparing their corresponding children. A Largest Common Subsequence (LCS) matching can be used for this step. The Object-Compare methods allows efficient, robust and optimal comparison of objects at the same complexity of the Ordered Tree comparisons.

To handle the extra freedom in the expressiveness of DAGs, we use a DAG-Compare algorithm, see "Lin, et al., "Coding and Comparison of DAGs as a novel neural structure with application to on-line handwritten recognition," IEEE Trans Signal Processing, 1996, incorporated herein by reference. We find the two best-matching paths between two DAGs. Although more general, the DAG-Compare is of the same order complexity as the LCS search. Lastly, we should mention that the constraints on the containment hierarchy (as a DAG) allow us to use the POT-Compare algorithm, but the POT is merely a subset of our generic content entity.

Applications

The content description scheme described above is not only an expressive means of describing content entities, but also provides a robust similarity measure that is computationally efficient and can seamlessly integrate various descriptions, both semantic and syntactic. Within the description scheme according to our invention, content entities, their attributes and their relations form a basic hyperlink network such as available from the HTTP standard.

By constraining our graph structures to Directed Acyclic Graphs onto the hierarchy of our content entities and their descriptions, we can give an extra expressiveness over ordered trees while maintaining computational complexity for robust comparison between content entities that is equivalent to an ordered tree comparison.

Freedom in Expressiveness

There is no strict hierarchy of content entities: any object may strictly contain another object as long as the containment is not contradictory. Instead of a tree hierarchy, the containment relation over the content entities induces a directed acyclic graph. Acircularity is maintained by disallowing contradictory containment. The restriction on cycles enables an efficient recursive formulation of comparison.

Focusing on the DAG structure, we map the DAG structure of DAG-Composition as follows: edges represent content entities, and nodes correspond to breakpoints segmentation. We can structure the object as a configuration of contained content entities within DAGs according to a predefined topological order. The restrictions on the DAGs compared to a general graph structure is its topological ordering. This order may be temporal or spatial, but it must be 1D. By following the order and obeying connectivity, a subgraph of DAG structure leads to a new concept: an ordered path represents a particular interpretation of

multimedia content, i.e. a representative view of the content entity as an ordered subset of its contained entities.

Because a DAG can contain multiple ordered paths, the DAG becomes a compact representation of the multiple interpretations of the data. The DAG data structure allows for the concept of parallel paths; thus, the DAG may integrate both semantic and syntactic elements through this parallel structure. The semantic and syntactic elements are not necessary equivalent, but, within the context of the DAG structure, they can be made interchangeable by placing them on these parallel constructions and its ordering.

These functionalities are a subset of a generic graph structure. However, as most graph matching problems are still open, these restrictions will allow us to compare these expressive structures. Although this ordering constrains the expressiveness of a DAG-composition, it does allow for element alignment in robust comparison of content entities.

Universal Multimedia Access

Because our description scheme is capable of representing and comparing multiple interpretations of multimedia content, it fits very well with the concept of Universal Multimedia Access (UMA). The basic idea of UMA, as shown in Figure 6, is to enable client devices 601 with limited communication,

processing, storage and display capabilities to access, via a network 602, rich multimedia content 603 maintained by a server device 604.

Recently, several solutions have focussed on adapting the multimedia content to the client devices. UMA can be provided in two basic ways - the first by storing, managing, selecting, and delivering different versions of the media objects (images, video, audio, graphics and text) that comprise the multimedia presentations. The second way is by manipulating the media objects on-the-fly, such as by using methods for text-to-speech translation, image and video transcoding, media conversion and summarization. This allows the multimedia content delivery to adapt to the wide diversity of client device capabilities in communication, processing, storage, and display.

Our description scheme can support UMA through the first item mentioned above, that is, depending on the client-side capabilities, the server-side may choose to send a more elaborate interpretation of the multimedia content or simply send a brief summary of the multimedia content. In this way, our description scheme acts as a managing structure that helps decide which interpretation of the multimedia content is best suited for the client-side devices. As part of the attributes for a content entity, the requirements may include items such as the size of each image or video frame, the number of video frames in the multimedia content, and other fields that pertain to resource requirements.

Ranking

As an additional feature, the content entities in a DAG can have associated ranks. Figure 7 shows a DAG 700 including edges 701-709 having associated ranks (R) 711-719. The ranking is according to the attributes, e.g., semantic intensity, syntactic direction, spatial, temporal, and so forth. The ranking can be in an increasing or decreasing order depending on some predetermined scale, for example a scale of one to ten, or alternatively, ten to one.

For example, the various segments of an "adventure-action" movie video can be ranked on a scale of 1-10 as to the intensity of the "action" in the movie. Similarly, the segments of a sports video, such as a football match can be ranked, where a scoring opportunity receives a relatively high score, and an "injury" on the field receives a relatively low score. Segments of gothic romance videos can be ranked on the relative level of "romantic" activity, horror films on the level of the levels of fright inducing scenes, comedies on their level of humor, rock videos on their loudness, and so forth. It should be understood that the measurements can be based on the semantic and/or the syntactic properties of the content.

The ranking can be manual, or machine generated. For example, a high number of short segments in a row would generally be indicative of a high level of activity, whereas long segments would tend to include a low level of activity. See Yeo et al., in "Rapid Scene Analysis on Compressed Video," IEEE

Transactions on Circuits and Systems for Video Technology, Vol. 5, No. 6, December 1995, page 533-544, for one way of measuring content attributes.

Once the various segments have been ranked, as shown in Figure 8, it becomes possible to traverse the DAG 800 according to the rank-ordering. The traversal can be considered a permuation of the content. In Figure 8, the arrows 801 indicates "skips," and the bolded edges indicate the only segments that are traversed. For example, here the ranking is based an "action," and only segments having an "action" ranking of eight or greater are traversed. It should be apparent that the traversing can be according to other rank orderings of the content.

Summary

Specifying a particular rank-based traversal in effect allows one to summarize a video. The "summary" shown in Figure 8 is a "high-action" summary. Thus, if summaries for two different video are extracted based on the same ranking criteria, the summaries can be compared with the scheme as shown in Figure 5. The advantage here is that when the videos are fairly lengthy, extraneous segments not germane to the comparison can be rapidly skipped and ignored to provide a more meaningful and faster comparison.

In another embodiment, as shown in Figure 9, some or all of the "primary content" entities 711-719 have associated secondary content entities (2^0) 901-

909. A secondary content entity characterizes its associated primary entity in a different manner. For example, a fifteen minute interview clip of a person speaking, can be associated with just one frame of the segment, a still image of the same person, or perhaps, text containing the persons name, and a brief description of what the person is saying. Now, a traversal can be via the primary or associated secondary content entities, and a summary can be the primary content entities, or the secondary content entities, or a mix of either. For example, a low bandwidth summary of a video would include only textual secondary entities in its traversal or selected permutations, and perhaps a few still images.

Although the invention has been described by way of examples of preferred embodiments, it is to be understood that various other adaptations and modifications may be made within the spirit and scope of the invention. Therefore, it is the object of the appended claims to cover all such variations and modifications as come within the true spirit and scope of the invention.

CLAIMS

We claim:

1. A method for ordering multimedia content, comprising the steps of:

segmenting the multimedia content to extract objects;

extracting and associating features of the objects to produce content entities;

coding the content entities to produce directed acyclic graphs of the content entities, each directed acyclic graph representing a particular interpretation of the multimedia content;

measuring attributes of each content entity; and

assigning the measured attributes to each corresponding content entity in the directed acyclic graphs to rank order the multimedia content.

2. The method of claim 1 wherein the measured attributes include intensity attributes.

3. The method of claim 1 wherein the measured attributes include direction attributes.

4. The method of claim 1 wherein the measured attributes include spatial attributes.

5. The method of claim 1 wherein the measured attributes include temporal attributes.

6. The method of claim 1 wherein the measured attributes are arranged in an increasing rank order.

7. The method of claim 1 wherein the measured attributes are arranged in an decreasing rank order.

8. The method of claim 1 further comprising the step of:

traversing the multimedia content according to the directed acylic graph and the measured attributes assigned to the content entities.

9. The method of claim 1 further comprising the step of:

summarizing the multimedia content according to the directed acylic graph and the measured attributes assigned to the content entities.

10. The method of claim 1 wherein the multimedia content is a three dimensional video sequence.

11. The method of claim 1 wherein nodes of the directed acyclic graphs represent the content entities and edges represent breaks in the segmentation, and the measured attributes are associated with the corresponding edges.

12. The method of claim 8 wherein at least one secondary content entity is associated with a particular content entity, and wherein the secondary content entity is selected during the traversing.

13. The method of claim 9 wherein a summary of the multimedia is a selected permutation of the content entities according to the associated ranks.

ABSTRACT OF THE DISCLOSURE

A method for generating a representation of multimedia content by first segmenting the multimedia content spatially and temporally to extract objects. Feature extraction is applied to the objects to produce semantic and syntactic attributes, relations, and a containment set of content entities. The content entities are coded to produce directed acyclic graphs of the content entities, where each directed acyclic graph represents a particular interpretation of the multimedia content. Attributes of each content entity are measured and the measured attributes are assigned to each corresponding content entity in the directed acyclic graphs to rank order the multimedia content.

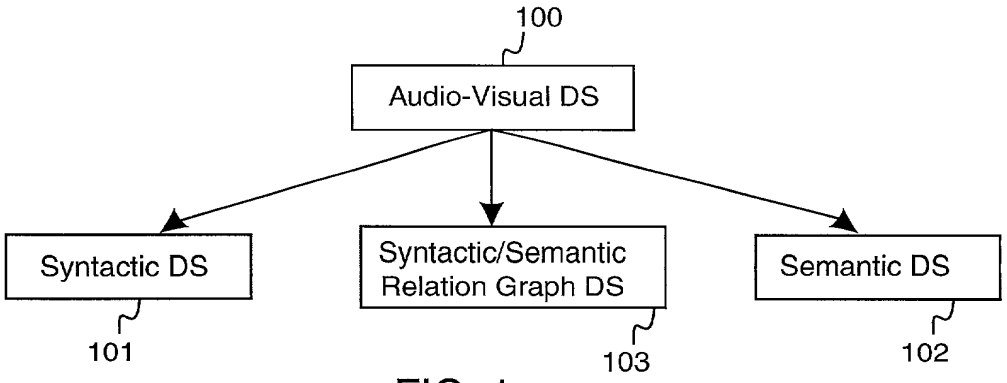


FIG. 1a

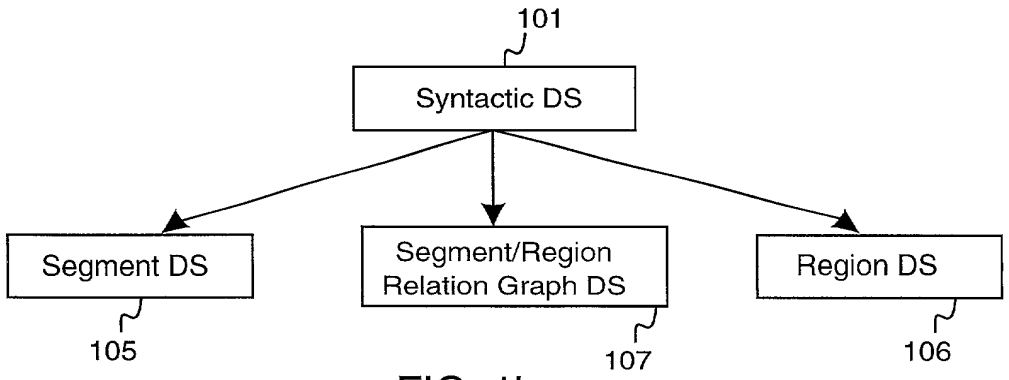


FIG. 1b

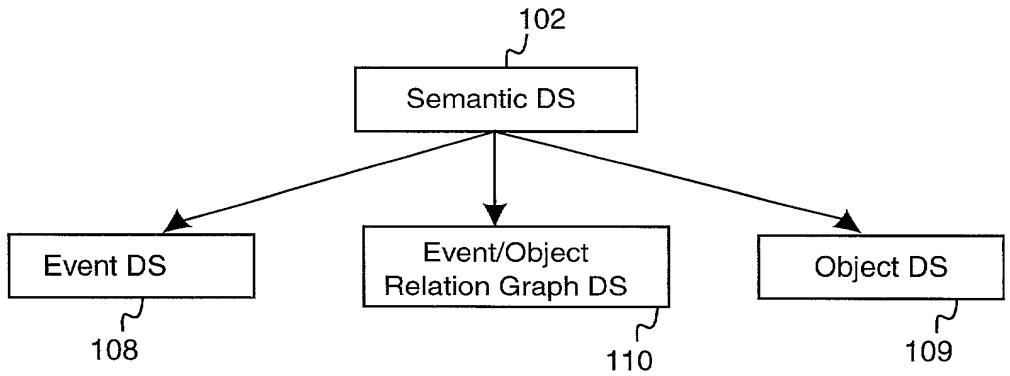


FIG. 1c

Prior Art

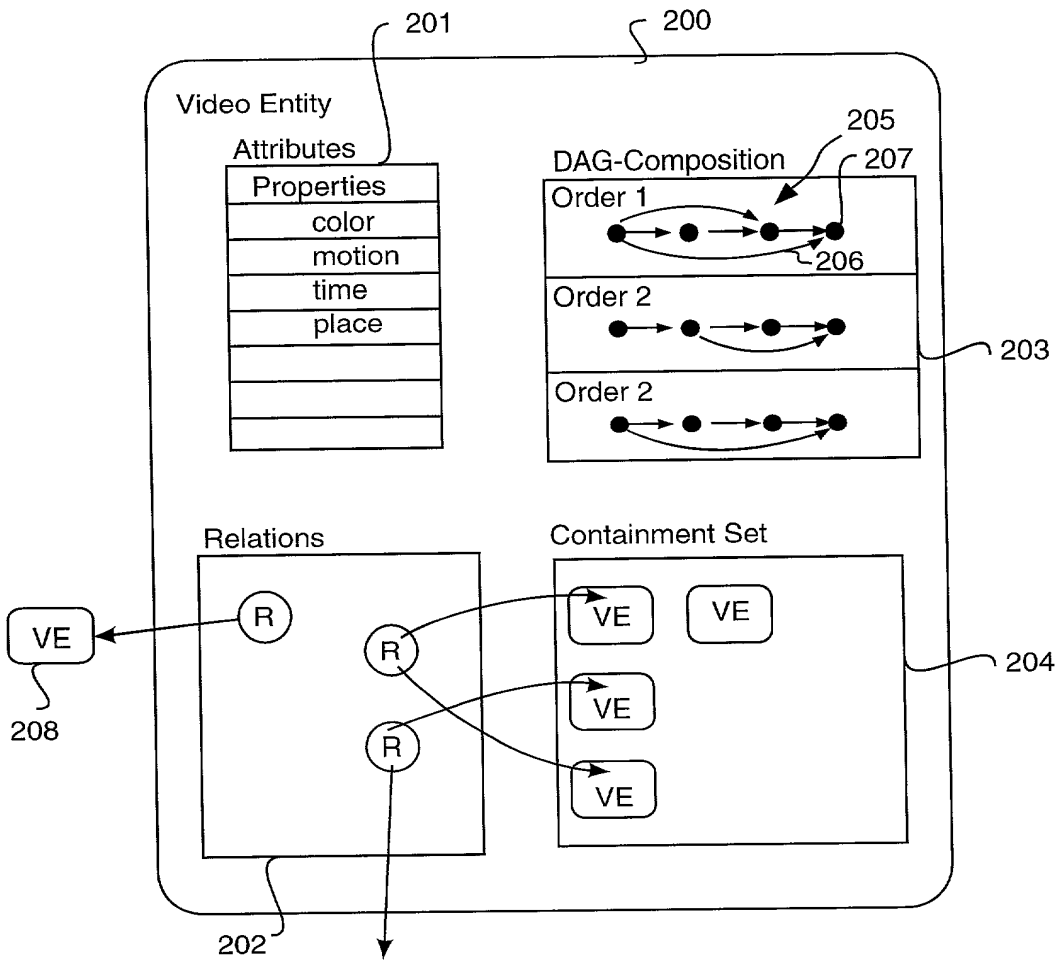

201
200
Video Entity
Attributes
Properties
color
motion
time
place
205
DAG-Composition
207
Order 1
206
Order 2
203
Order 2
Relations
Containment Set
VE
R
R
R
VE
VE
VE
VE
204
208
202

FIG. 2

Docket No. 2005-0158

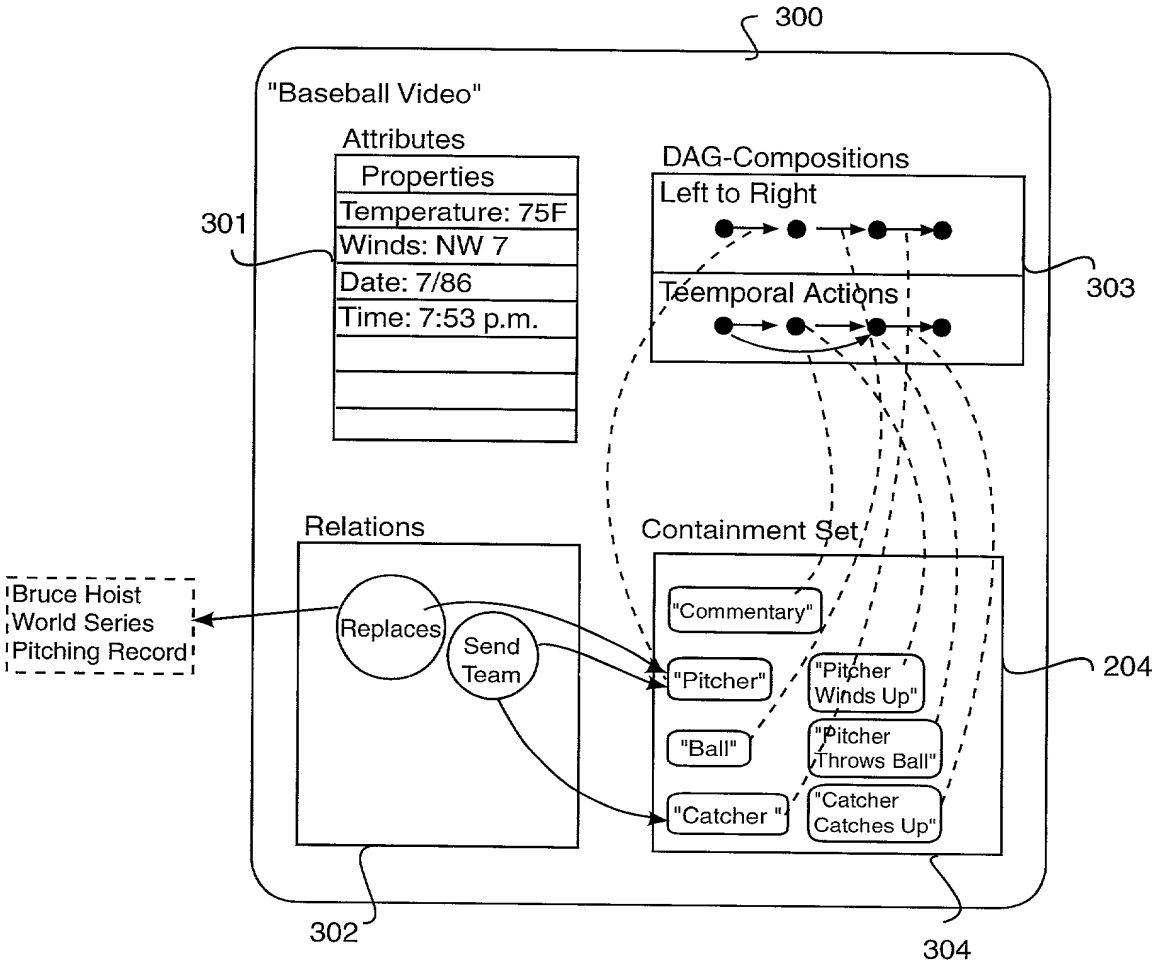


FIG. 3a

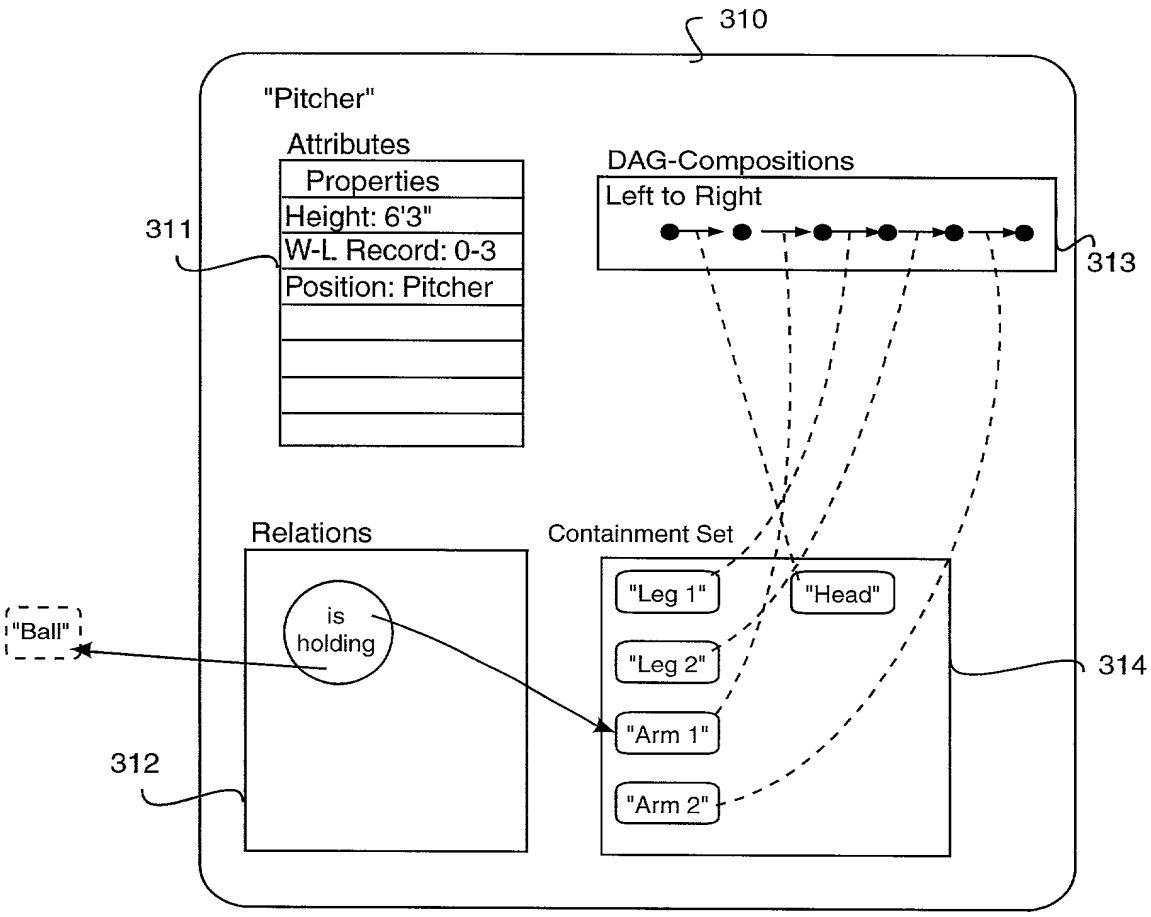

310
"Pitcher"
Attributes
Properties
Height: 6'3"
W-L Record: 0-3
Position: Pitcher
311
DAG-Compositions
Left to Right
313
Relations
is holding
"Ball"
312
Containment Set
"Leg 1"
"Head"
"Leg 2"
"Arm 1"
"Arm 2"
314

FIG. 3B

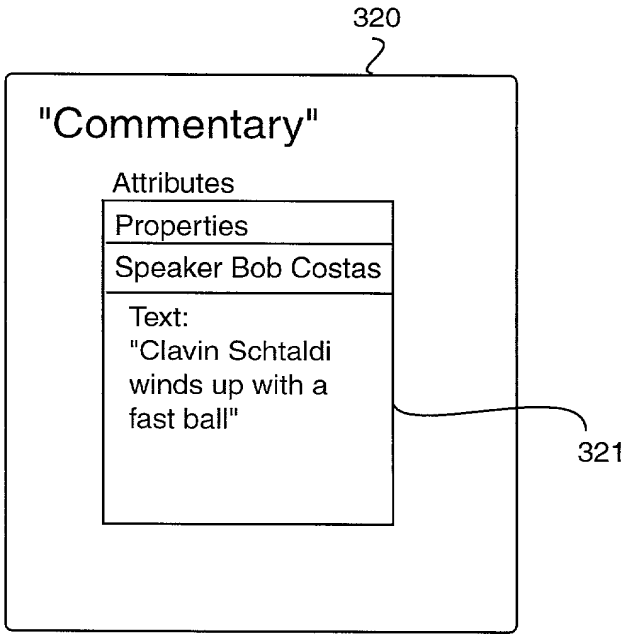

320
"Commentary"
Attributes
Properties
Speaker Bob Costas
Text:
"Clavin Schtaldi winds up with a fast ball"
321

FIG. 3c

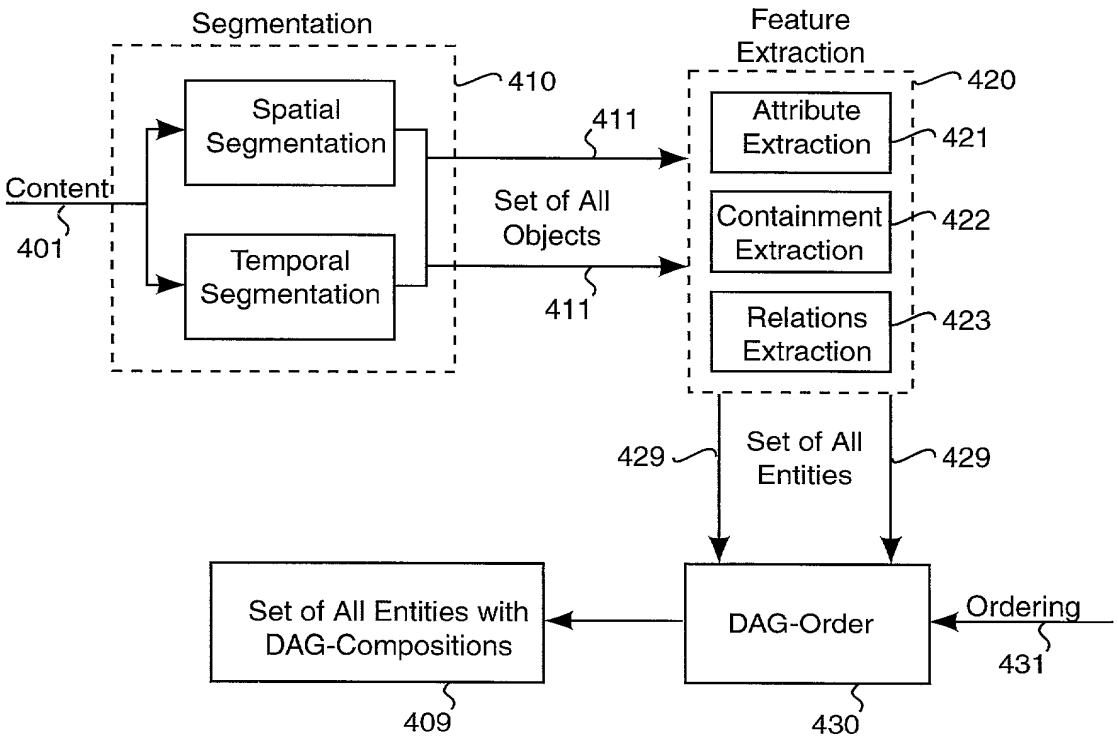

Segmentation
Feature Extraction
410
420
Spatial Segmentation
Attribute Extraction
421
411
Content
401
Set of All Objects
Containment Extraction
422
Temporal Segmentation
411
Relations Extraction
423
429
Set of All Entities
429
Set of All Entities with DAG-Compositions
DAG-Order
Ordering
431
409
430

FIG. 4

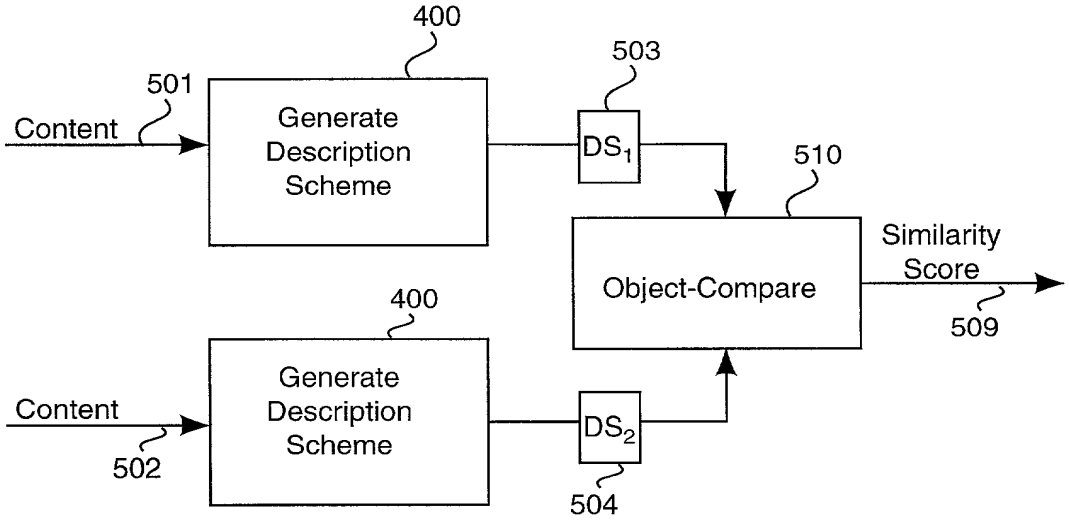

400
503
501
Content
Generate Description Scheme
DS1
510
Object-Compare
Similarity Score
509
400
Content
502
Generate Description Scheme
DS2
504

FIG. 5

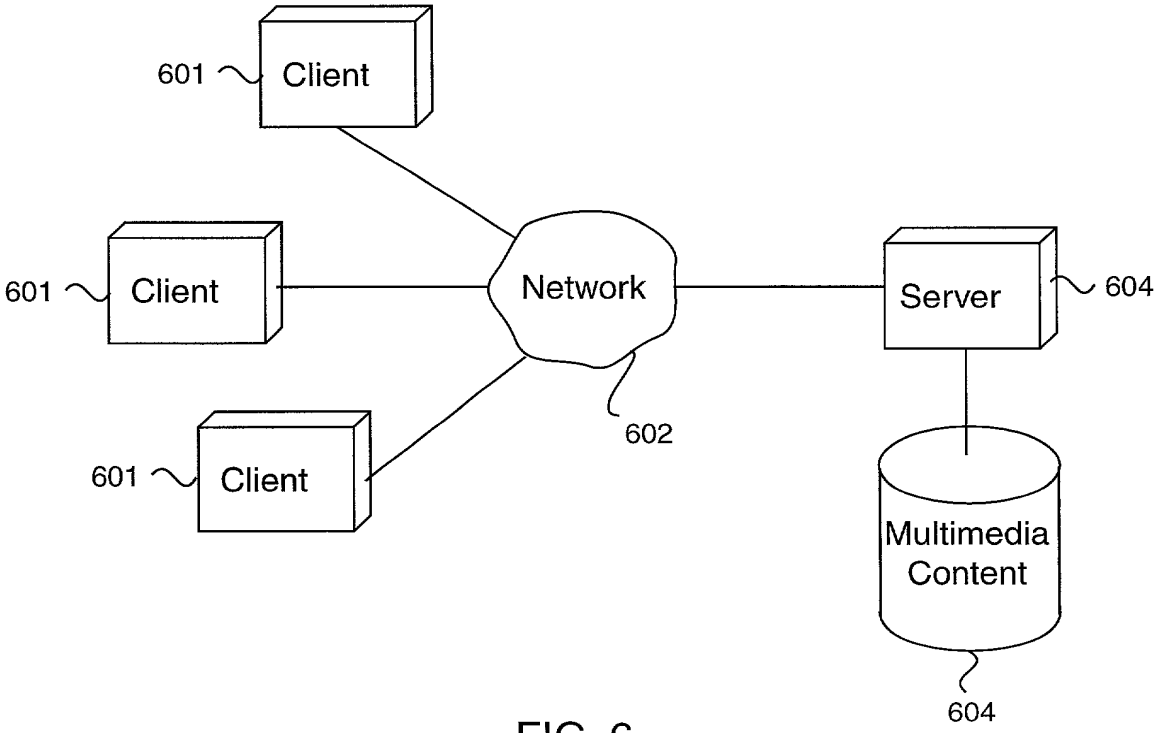

601
Client
601
Client
Network
Server
604
601
Client
602
Multimedia Content
604

FIG. 6

05/05/2006, 2:20:55 PM

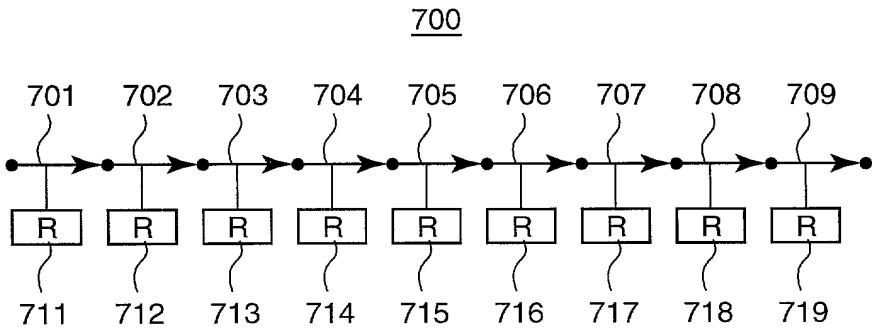


FIG. 7

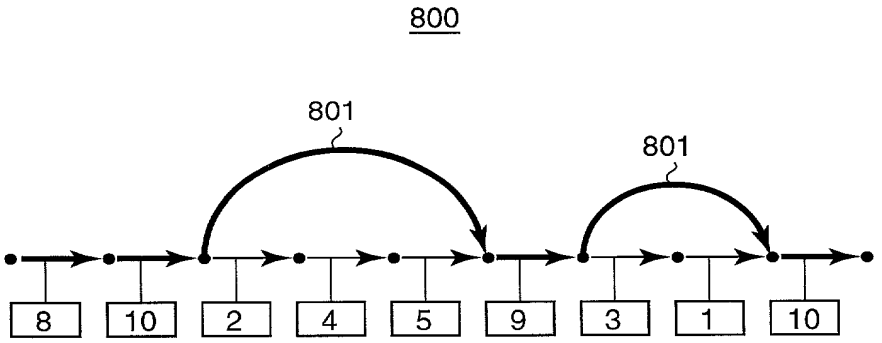


FIG. 8

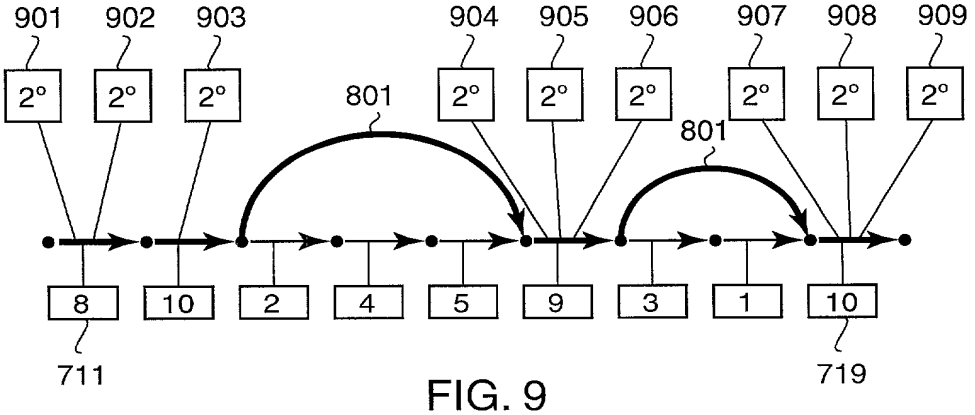


FIG. 9

DECLARATION AND POWER OF ATTORNEY

DECLARATION:

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe, the below named inventors are the original, first and joint inventors of the subject matter which is claimed and for which a patent is sought on the invention for METHOD FOR REPRESENTING AND COMPARING MULTIMEDIA CONTENT ACCORDING TO RANK, the specification of which is attached hereto unless the following box is checked.

[_] was filed on _>_ as Application Serial Number _>_ and was amended on _>_ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I acknowledge the duty to disclose information which is material to patentability in accordance with Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119(a)-(d) of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

PRIOR FOREIGN APPLICATIONS

Number	Country	Date Filed	Priority Claimed (Yes/No)
>	>	>	>
>	>	>	>
>	>	>	>

I hereby claim the benefit under Title 35, United States Code §119(e) of any United States Provisional application(s) listed below.

APPLICATION NUMBER	FILING DATE
>	>
>	>

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

PRIOR UNITED STATES APPLICATIONS

Application Serial Number	Filing Date	Status
>	>	>
>	>	>
>	>	>

I hereby declare that all statements made of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

POWER OF ATTORNEY:

On behalf of Mitsubishi Electric Information Technology Center America, Inc., Assignee of my entire right, title and interest, I hereby appoint the following attorney with full power of substitution to act exclusively for Mitsubishi Electric to prosecute this application and transact all business in the Patent and Trademark Office connected therewith: Dirk Brinkman, Reg. No. 35,460.

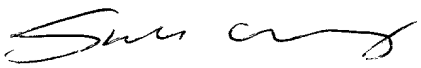
All correspondence should be addressed to:

Patent Department
Mitsubishi Electric Information Technology
Center America, Inc.
201 Broadway
Cambridge, Massachusetts 02139

All telephone calls should be directed to Dirk Brinkman, telephone number (617) 621-7539.

Inventor's Full Name:	Ajay Divakaran
Inventor's Signature:	Ajay Divakaran
Date:	March 3, 2000
Residence: (City, State and/or country)	20, Cobb Road, Denville, NJ 07834
Citizenship:	India
Post Office Address:	20, Cobb Road, Denville, NJ 07834

Inventor's Full Name:	Anthony Vetro
Inventor's Signature:	Anthony Vetro
Date:	March 3, 2000
Residence: (City, State and/or country)	113 Regis Drive Staten Island, New York 10314
Citizenship:	United States of America
Post Office Address:	113 Regis Drive Staten Island, New York 10314

Inventor's Full Name:	Huifang Sun
Inventor's Signature:	
Date:	March 3, 2000
Residence: (City, State and/or country)	61 Kinglet Drive South Cranbury, New Jersey 08572
Citizenship:	United States of America
Post Office Address:	61 Kinglet Drive South Cranbury, New Jersey 08572

Docket #: ATL-042